STATS 4CI3/6CI3 Winter 2021

ASSIGNMENT 5

Submit to Crowdmark using the link that was emailed to you.

Due before 11 PM on Friday, March 26th.

Your assignment must conform to the Assignment Standards listed below.

Assignments submitted up to 24 hours late will incur a 30% penalty. Assignments submitted more than 24 hours late will receive a zero grade.

Answer all questions, stating your answer and showing your code. Not all questions carry equal marks.

Set a seed where appropriate to make your work reproducible

1. **(20 MARKS)**

   Suppose we have the following data from a study set up to determine whether or not regular doses of a certain quantity of aspirin were effective at preventing heart attacks.

Treatment	Heart Attacks	Subjects
Aspirin	98	11037
Placebo	195	11034

   (a) Using a bootstrap estimate of size 1000, calculate an estimate of the ratio of the rates of heart attack for the two treatment groups.

   (b) Produce a 95% bootstrap percentile interval for your estimate.

   (c) Produce a histogram showing the distribution of your estimate, as well as the 95% bootstrap percentile interval.

   (d) Is there evidence that aspirin is preventing heart attacks? Justify your answer.

2. **(30 MARKS)**

   Presume you want to estimate the coefficient of variation, CV $= \sqrt{Var}/\bar{x}$, for the following observations:

5.32, 9.53, 7.44, 5.71, 6.85, 8.63, 5.98, 6.19, 5.2,
6.81, 8.74, 7.22, 9.22, 6, 6.5, 4.18, 5.12, 7.21, 6.52, 7.31,
12.8, 5.86, 6.82, 6.86, 7.48

(a) Calculate the coefficient of variation (CV) directly, using the equation stated above.

(b) Compute the CV using a single bootstrap sample.

(c) Write code to compute the CV using 1000 bootstrap samples, and then

i. Calculate the mean and variance of this collection of bootstrap samples.

ii. Produce a histogram to illustrate the distribution of your estimate for the CV.

iii. Calculate a 95% bootstrap percentile interval for your estimate.

iv. Calculate the bias of your estimate.

3. **(27 MARKS)**

For the same set of observations as in Question 2, use the Jack-knife method to

(a) Calculate an estimate of the CV.

(b) Calculate the variance of this estimate.

(c) Produce a histogram to illustrate the distribution of your estimate for the CV.

(d) Calculate a 95% bootstrap percentile interval for your estimate.

(e) Calculate the bias of your estimate.

4. **(20 MARKS)**

(a) With regard to the basic Bayesian statistics process, describe how a Posterior Distribution of a parameter θ is created by using the observed data, i.e. x, $\pi(\theta \mid x)$.

(b) Outline the Monte Carlo Bayesian Inference technique.

(c) Given a sample from the posterior distribution, give a Monte Carlo estimater of $\widehat{Var}(\theta \mid x_1, ..., x_n)$ for the posterior distribution.

(d) Define a $100(1-\alpha)\%$ Bayesian Credible Interval and briefly outline how you might find such an interval.

(e) What is the defining characteristic of a non-informative prior?

5. **(3 MARKS)** With regard to the ggmcmc package in R, define (i.e. provide a "Description") of the following functions:

   (a) ggs_density

   (b) ggs_compare_partia

   (c) ggs_traceplot

   (d) ggs_running

   (e) ggs_autocorrelation

   (f) ggs_caterpillar

Assignment Standards

- LaTeX is strongly recommended but not strictly required. The use of Markdown in R studio is also recommended.
- Submit your assignment as one **.pdf document**. **All R code should be included inline.**
- Do not include a title page. The title, your **name and student number** should be printed at the top of the first page.
- The writing and referencing should be appropriate to the university level.
- Various tools, including publicly available internet tools, may be used by the instructor to check the originality of submitted work.
